POLITECNICO DI TORINO
Repository ISTITUZIONALE

Set membership fault detection for nonlinear dynamic systems

Original
Set membership fault detection for nonlinear dynamic systems / Karimshoushtari, Milad; Spagnolo, Luigi; Novara, Carlo - In: Data-Driven Modeling, Filtering and Control: Methods and Applications[s.l] : IET the Institution of Engineering and Technology, 2019. - ISBN 9781785617133. - pp. 239-264 [10.1049/PBCE123E_ch12]

Availability:
This version is available at: 11583/2831720 since: 2020-06-03T15:55:48Z

Publisher:
IET the Institution of Engineering and Technology

Published
DOI:10.1049/PBCE123E_ch12

Terms of use:
openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

05 May 2023

Chapter 1

Set Membership Fault Detection for Nonlinear Dynamic Systems

Milad Karimshoushtari, Luigi Spagnolo, Carlo Novara[1]

In this chapter, an innovative approach to fault detection for nonlinear dynamic systems is proposed, based on the recently introduced quasi-local Set Membership identification method, overcoming some relevant issues proper of the "classical" techniques. The approach is based on the direct identification from experimental data of a suitable filter and related uncertainty bounds. These bounds are used to detect when a change (e.g., a fault) has occurred in the dynamics of the system of interest. The main advantage of the approach compared to the existing methods is that it avoids the utilization of complex modeling and filter design procedures, since the filter/observer is directly designed from data. Other advantages are that the approach does not require to choose any threshold (as typically done in many "classical" techniques) and it is not affected by under-modeling problems. An experimental study regarding fault detection for a drone actuator is finally presented to demonstrate the effectiveness of the proposed approach.

1.1 Introduction

Consider a discrete-time nonlinear system in state-space form:

$$\begin{aligned} z_k &= f^o\left(z_{k-1}, u_{k-1}\right) + d_k \\ y_k &= z_{i,k} \end{aligned} \tag{1.1}$$

where $z_k \in \mathbb{R}^{n_z}$ is the state, $y_k \in \mathbb{R}$ is the output, $z_{i,k}$ is a component of z_k, $u_k \in \mathbb{R}^{n_u}$ is the input, $d_k \in \mathbb{R}^{n_d}$ is a bounded disturbance and $k = 0, 1, 2, \ldots$ is the discrete time index. Assume that the input u_k and the state z_k are measured. Note that the assumption of measuring the state is not strictly necessary: the fault detection approach proposed in the following can be applied with minor modifications using an input-output system representation, see Remark 1 below.

A "classical" approach to fault detection is to identify a model of the system (1.1) and to design a filter/observer on the basis of the identified model. The designed filter/observer is then used to generate on-line a suitable residual signal. The fault

[1]The Authors are with Politecnico di Torino, Italy. E-mail: milad.karimshoushtari@polito.it, luigi.spagnolo@polito.it, carlo.novara@polito.it.

is detected when the residual exceeds a given threshold, see e.g. [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. However, the design of the filter/observer may be hard in the presence of nonlinear and/or uncertain dynamics. Indeed, designing an optimal filter from a nonlinear model is in general not possible, and approximate filters only, such as the Extended Kalman filters, can be actually obtained. These kind of filters may often be inaccurate and not even guarantee the estimation error stability. Moreover, the choice of the threshold may be critical, especially when poor prior information on the system is available. Another relevant issue is that, in real-world applications, the system (1.1) is unknown and only approximate models can be identified from finite data; evaluating the effects of the modeling error on the estimation error of the filter designed from the approximated model is a largely open problem.

Set Membership fault detection methods have been introduced to efficiently deal with modeling errors, [11, 12, 13, 14, 15, 16, 17]. These methods have been mainly developed for linear systems, while only few of them deal with nonlinear systems, [15, 16, 17]. Typically, in Set Membership methods, a suitable estimation interval is computed online and the fault is detected when one or more measured variables fall outside this interval.

In this chapter, following this Set Membership philosophy, an innovative approach to fault detection is considered, allowing us to overcome the above issues. The main advantage of this approach compared to the existing methods ("classical" and Set Membership) is that it avoids the utilization of difficult filter design procedures, since the filter/observer is directly designed from data. Other advantages with respect to the "classical" methods are that the approach does not require to choose any threshold and it is not affected by modeling errors since no model is used. A further interesting feature is that the approach is computationally simple, in both the design and implementation phases.

The method proposed in this chapter represents an improvement with respect to the one of [18]. Indeed, in [18], filter design is performed by means of the so-called local nonlinear Set membership identification method, where the filter is obtained in the form of a linear combination of given basis functions. In this chapter, a so-called quasi-local method is presented, where no filter parametric form needs to be assumed. The filter is obtained directly from experimental data in a non-parametric closed form, thus not requiring the choice of a suitable set of basis functions. Such a quasi-local approach is similar to the so-called global approach of [19] but leads to the derivation of significantly less conservative uncertainty bounds.

The chapter is organized as follows. In section 1.2 the fault detection problem is formulated in the set membership framework, defining the types of assumptions and optimality concepts considered. In sections 1.3 and 1.4, the nonlinear set membership identification method of [19, 18] is summarized. Two versions of this method are discussed: the global one, where the filter is obtained in closed-form, and the local one, where the filter is obtained in the form of a linear combination of given basis functions, solving a convex optimization problem. In section 1.5, the quasi-local approach is presented. In section 1.6, two algorithms are presented to find the model parameters and an algorithm to build an adaptive set membership model. In section 1.7 a summary of the fault detection procedure is given. Finally, in section 1.8, the method is tested on a drone propeller in a real experimental setting.

1.2 Nonlinear Set Membership Fault Detection

Suppose that the function f^o in (1.1) is not know but a set of noise corrupted data is available, given by

$$\mathcal{D} = \{\widetilde{x}_k, \widetilde{y}_k\}_{k=1}^{L} \tag{1.2}$$

where $\widetilde{x}_k \doteq (\tilde{z}_{k-1}, \tilde{u}_{k-1})$. The proposed fault detection approach consist of the following main steps. First, a Set Membership filter for the system (1.1) is defined from the dataset (1.2) which gives us tight bounding functions $\overline{f}$ and $\underline{f}$, such that

$$\underline{f}\left(\widetilde{x}_k\right) \leqslant f_i^o\left(\widetilde{x}_k\right) \leqslant \overline{f}\left(\widetilde{x}_k\right), \, \forall k.$$

Then, a fault detection system F is defined, on the basis of the bounding functions $\overline{f}$ and $\underline{f}$. The inputs of F is $\widetilde{x}_k \doteq (\tilde{z}_{k-1}, \tilde{u}_{k-1})$, the outputs are the following:

$$\left.\begin{array}{l} \overline{y}_k \doteq \overline{f}\left(\widetilde{x}_k\right) + \varepsilon_k \\ \underline{y}_k \doteq \underline{f}\left(\widetilde{x}_k\right) - \varepsilon_k \end{array}\right\} k > L. \tag{1.3}$$

where ε_k is a bound on the noise d_k affecting the system. It will be shown in the following sections how to construct the functions $\overline{f}$ and $\underline{f}$ and to properly choose the involved parameters (e.g., ε_k). The rationale behind this fault detection scheme can be explained as follows.

Since $\tilde{y}_k = f_i^o\left(\widetilde{x}_k\right) + d_k$, we have that $\tilde{y}_k \leqslant \overline{y}_k$, $\forall k$. Similarly, it holds that $\tilde{y}_k \geqslant \underline{y}_k$, $\forall k$. It follows that $\tilde{y}_k > \overline{y}_k$ or $\tilde{y}_k < \underline{y}_k$ only if the function f^o has changed, i.e. only if some structural change has occurred in the system (1.1). On the basis of this result, fault detection is performed by checking on-line if $\tilde{y}_k > \overline{y}_k$ or $\tilde{y}_k < \underline{y}_k$: a fault is detected as soon as one of these two inequalities is satisfied.

Remark 1. *If the system state is not measured, the following input-output representation can be considered:*

$$\begin{array}{l} y_k = f^o\left(x_k\right) + d_k \\ x_k = \left(y_{k-1}, \ldots, y_{k-n_y}, u_{k-1}, \ldots, u_{k-n_u}\right) \end{array} \tag{1.4}$$

where $y_k \in \mathbb{R}^{n_y}$ is the measured output, $u_k \in \mathbb{R}^{n_u}$ is the measured input and $d_k \in \mathbb{R}^{n_d}$ is an unknown bounded disturbance. Note that the function f^o in (1.4) is different from the one in (1.1). The proposed fault detection approach can be applied considering this representation without significant modifications.

Remark 2. *The proposed fault detection approach can be applied to each state component (or output component, in the case where the input-output representation (1.4) is used), in order to obtain a multi-dimensional fault detection system, allowing us to improve the detection performance with respect to a mono-dimensional case.*

1.2.1 Problem Formulation

Consider a nonlinear function f^o defined by

$$y = f^o(x) \tag{1.5}$$

where $x \in \mathcal{X} \subset \mathbb{R}^{n_x}$, $\mathcal{X}$ is a compact connected set and $y \in \mathbb{R}$. Suppose that f^o is not known but a set of noise-corrupted data $\mathcal{D} = \{\widetilde{x}_k, \widetilde{y}_k\}_{k=1}^{L}$ is available, described by

$$\widetilde{y}_k = f^o(\widetilde{x}_k) + d_k, \quad k = 1, 2, \ldots, L \tag{1.6}$$

where d_k is a noise. Assume that the noise sequence $d = (d_1, d_2, \ldots, d_L)$ is unknown but bounded:

$$\|d\|_q \leqslant \mu \tag{1.7}$$

where $\|\cdot\|_q$ is the vector ℓ_q norm. A general formulation is developed in the following, allowing us to deal with the most important cases (i.e. $q = 2, \infty$) in a unified framework. As well known, the ℓ_2 norm is related to the energy of the considered signal, while the ℓ_∞ norm to its amplitude. The choice of this norm can be carried out on the basis of the prior knowledge on the energy or amplitude of the involved noise (if available) or by means of a trial and error procedure.

In the following sections, the problem of deriving from the available data an approximation $\hat{f}$ of f^o and evaluating tight estimate bounds on f^o is considered. The approximation is required to be accurate on the whole domain $\mathcal{X}$. The accuracy is measured by means of the following approximation error:

$$e(\hat{f}) \doteq \|f^o - \hat{f}\|_p \tag{1.8}$$

where $\|\cdot\|_p$ is a functional L_p norm, defined as

$$\|f\|_p \equiv \|f(\cdot)\|_p \doteq \begin{cases} [\int_{\mathcal{X}} |f(x)|^p \, dx]^{1/p}, p \in [1, \infty) \\ \text{ess sup}_{x \in \mathcal{X}} |f(x)|, p = \infty. \end{cases} \tag{1.9}$$

For a given estimate $\widehat{f}$, the related L_p error is given by (1.8). This error cannot be exactly computed since f^o is not known. It is only known that $f^o \in FFS$, where FFS is called the *Feasible Function Set*, that is, the set of all possible functions consistent with the available prior information and measured data. The formal definition of FFS will be given in the next sections, for three relevant specific cases. This motivates the following definition of identification error, often indicated as worst-case or guaranteed error.

Definition 1. Worst-case approximation error *of* $\widehat{f}$*:*

$$EN(\widehat{f}) \doteq \sup_{f \in FFS} \|f - \widehat{f}\|_p. \qquad \square \tag{1.10}$$

An optimal approximation is defined as a function f^{op} which minimizes the worst-case approximation error.

Definition 2. *An approximation* f^{op} *is* optimal *if*

$$EN(f^{op}) = \inf_{\widehat{f}} EN(\widehat{f}) \doteq \mathcal{R}_{\mathcal{I}}.$$

$\mathcal{R}_\mathcal{I}$ is called the radius of information *and is the minimum worst-case error that can be achieved on the basis of the available prior and experimental information.* □

Finding optimal approximations may be hard or not convenient, and sub-optimal solutions can be looked for. In particular, approximations called interpolatory are often considered in the literature, see e.g. [20], [21].

Definition 3. *An approximation f^I is* interpolatory *if*

$$f^I \in FFS. \qquad \square$$

A fundamental property of an interpolatory approximation is that it guarantees a worst-case error degradation of at most 2, [20], [21]. An approximation with this property is called almost-optimal.

Definition 4. *An approximation f^{ao} is* almost-optimal *if*

$$EN\left(f^{ao}\right) \leqslant 2 \inf_{\widehat{f}} EN(\widehat{f}). \qquad \square$$

In this chapter, the following problem is considered.

Problem 1. *From the data set* (1.2)*, find an approximation $\hat{f}$ of f^o*
(i) optimal or almost-optimal;
(ii) equipped with tight interval estimates $\overline{f}, \underline{f}$ for f^o. □

Remark 3. *In the SM and approximation theory literature, two optimality concepts are typically considered: local and global optimality, [20, 19]. The worst-case error (1.10) is a local error, since it depends also on the function f^o and the data $\mathcal{D}$, i.e. $EN(\widehat{f}) = EN(\widehat{f}, f^o, \mathcal{D})$. A global identification error is also often considered, defined as:*

$$EN^g(\widehat{f}) \doteq \sup_{\substack{f^o \in \mathcal{F}(\Gamma) \\ \mathcal{D} \in \{d : \|d\|_q \leqslant \mu\}}} EN(\widehat{f}, f^o, \mathcal{D}).$$

An approximation f^g is called globally optimal if $EN^g(f^g) = \inf_{\widehat{f}} EN^g(\widehat{f})$. This is the optimality concept usually investigated in the SM context and approximation theory literature, [20]. Note that a locally optimal algorithm f^{op} is globally optimal, but f^g is not in general locally optimal. Therefore, the local optimality concept investigated in this chapter is stronger and thus less conservative than the global optimality concept investigated in the above mentioned literature.

1.3 Nonlinear Set Membership Identification: Global Approach

In order to ensure a bound on the approximation error (1.8), some assumptions have to be made on the noise affecting the data and on the unknown function f^o. In this chapter, the noise sequence $d = (d_1, d_2, \ldots, d_L)$ in (1.6) is assumed to be bounded according to (1.7). Differently from [22], where f^o is assumed to be parametrized by a finite set of basis functions, a mild regularity assumption is made here on f^o, not requiring any knowledge on its parametric form. In particular, we assume that the function f^o is Lipschitz continuous on $\mathcal{X}$:

$$f^o \in \mathcal{F}(\Gamma) \tag{1.11}$$

for some $\Gamma < \infty$, where

$$\mathcal{F}(\Gamma) \doteq \{f : |f(x) - f(\widehat{x})| \leqslant \Gamma \|x - \widehat{x}\|_2, \forall x, \widehat{x} \in \mathcal{X}\}.$$

This allows us to introduce the *Feasible Function Set* (FFS).

Definition 5. *The Feasible Function Set is*

$$FFS \doteq \left\{ f \in \mathcal{F} : \|\widetilde{y} - f(\widetilde{x})\|_q \leqslant \mu \right\} \tag{1.12}$$

where $\widetilde{y} = (\widetilde{y}_1, \ldots, \widetilde{y}_L)$ *and* $f(\widetilde{x}) \doteq (f(\widetilde{x}_1), \ldots, f(\widetilde{x}_L))$. □

According to this definition, FFS is the set of all functions consistent with the prior assumptions and data. If the prior assumptions hold, then $f^o \in FFS$, that is an important property for evaluating the accuracy of any estimate.

In the Set Membership framework, the validation of prior assumptions is a fundamental step. It is usual to introduce the concept of prior assumption validation as consistency with the available data: the prior assumptions are considered validated if at least one estimate consistent with these assumptions and the data exists, i.e. if FFS is not empty, see e.g. [21] and [23].

Definition 6. *Prior assumptions are validated if* $FFS \neq \varnothing$ □

The following theorem gives a necessary and a sufficient condition for the validation of prior assumptions. Let us define the following functions:

$$\begin{aligned} \overline{f}(x) &\doteq \min_{k=1,\ldots,L} (\overline{h}_k + \Gamma \|x - \widetilde{x}_k\|_2) \\ \underline{f}(x) &\doteq \max_{k=1,\ldots,L} (\underline{h}_k - \Gamma \|x - \widetilde{x}_k\|_2) \end{aligned} \tag{1.13}$$

where $\overline{h}_k \doteq \widetilde{y}_k + \varepsilon_k$ and $\underline{h}_k \doteq \widetilde{y}_k - \varepsilon_k$.

Necessary and sufficient conditions for checking the assumptions validity are now given.

Theorem 1. *(i) A necessary condition for prior assumptions to be validated is:* $\overline{f}(\widetilde{x}_k) \geqslant \underline{h}_k$ *,* $\underline{f}(\widetilde{x}_k) \leqslant \overline{h}_k$*,* $k = 1, \ldots, L$*.*
(ii) A sufficient condition for prior assumptions to be validated is: $\overline{f}(\widetilde{x}_k) > \underline{h}_k$ *,* $\underline{f}(\widetilde{x}_k) < \overline{h}_k$*,* $k = 1, \ldots, L$*.*

Proof. see [19]. □

Note that the fact that prior assumptions are validated, i.e., that they are consistent with the present data, does not exclude that they may be invalidated by future data. In the reminder of chapter, it is assumed that the sufficient condition holds. If not, values of the constants appearing in the assumptions on function f^o and on the noise d_k have to be suitably modified. The above validation Theorem can be used for assessing the values of such constants so that sufficient conditions holds.

Now let the function f_c be defined as

$$f^c(x) \doteq \frac{1}{2}[\underline{f}(x) + \overline{f}(x)] \tag{1.14}$$

where $\underline{f}(x)$ and $\overline{f}(x)$ are given in (1.13). The next result shows that the estimate f^c is optimal according to definition 2 for any L_p norm.

Theorem 2. *Assume that:*
(i) The noise affecting the measurements $\{\widetilde{x}_k, \widetilde{y}_k\}_{k=1}^{L}$ *is bounded according to (1.7).*
(ii) The function f^o *is Lipschitz continuous according to (1.11).*
Then, for $q = \infty$ *and for any* $p \in [1, \infty]$*:*
(i) The approximation f^c *defined in* (1.14) *is optimal.*
(ii) The worst-case approximation error of f^c *is given by*

$$E(f^c) = \frac{1}{2}\|\overline{f} - \underline{f}\|_p = \inf_{\widehat{f}} EN(\widehat{f}) = \mathcal{R}_{\mathcal{I}}. \tag{1.15}$$

Proof. see [19]. □

1.3.1 Interval estimates

The interval estimates $\overline{f}, \underline{f}$ of the unknown function f^o are now given for the cases where the noise is bounded in ℓ_∞ norm (i.e. $q = \infty$ in (1.7)).

Theorem 3. *Assume that:*
(i) The noise affecting the measurements $\{\widetilde{x}_k, \widetilde{y}_k\}_{k=1}^{L}$ *is bounded according to (1.7).*
(ii) The function f^o *is Lipschitz continuous according to (1.11).*
Then, for $q = \infty$ *and for any* $x \in \mathcal{X}$, $f^o(x)$ *is tightly bounded as*

$$\underline{f}(x) \leqslant f^o(x) \leqslant \overline{f}(x) \tag{1.16}$$

Proof. see [19]. □

$\overline{f}$ and $\underline{f}$ are also called *optimal bounds* since they are tightest upper and lower bounds of f^o.

1.4 Nonlinear Set Membership Identification : Local Approach

Suppose that a preliminary approximation f^* of the function f^o has been obtained using any method. This approximation is of the form

$$f^*(x) = \sum_{i=1}^{N} a_i \phi_i(x) \tag{1.17}$$

where $\phi_i : \mathcal{X} \to \mathbb{R}$ are Lipschitz continuous basis functions and $a_i \in \mathbb{R}$ are parameters identified by means of some suitable algorithm (two algorithms will be presented in Section 1.4.2). The choice of the basis functions ϕ_i is clearly an important step of the identification process, see e.g. [24, 25, 26]. In several cases of practical interest, the basis functions are known a priori to belong to some "large" set of functions, see e.g. the example presented in [22]. The sparse approximation algorithms presented below can be applied in these cases to select within this "large" set the functions which are important for providing an accurate description of the system under investigation. In other cases, the basis functions are not known a priori and their choice can be carried out considering the numerous options available in the literature (e.g. gaussian, sigmoidal, wavelet, polynomial, trigonometric). See [24] for a discussion on the main features of the most used basis functions and for indications for their choice.

Define the following *residue function*:

$$f_\Delta(x) \doteq f^o(x) - f^*(x). \tag{1.18}$$

From (1.11) and from the Lipschitz continuity of ϕ_i, it follows that f_Δ is Lipschitz continuous over the set $\mathcal{X}$:

$$f_\Delta \in \mathcal{F}(\Gamma_\Delta). \tag{1.19}$$

for some $\Gamma_\Delta < \infty$. Note that the Lipschitz constant Γ_Δ can be estimated by means of the algorithm presented in section 1.6.

Remark 4. *The inclusion* (1.19) *corresponds to assume a* global *maximum rate of variation for f_Δ but a* local *maximum rate of variation for f^o. Indeed, for every $x, \widehat{x} \in \mathcal{X}$, the following inequalities hold:*

$$-\Gamma_\Delta \leqslant \tfrac{f_\Delta(x)-f_\Delta(\widehat{x})}{\|x-\widehat{x}\|_2} \leqslant \Gamma_\Delta$$

$$\tfrac{f^*(x)-f^*(\widehat{x})}{\|x-\widehat{x}\|_2} - \Gamma_\Delta \leqslant \tfrac{f^o(x)-f^o(\widehat{x})}{\|x-\widehat{x}\|_2} \leqslant \tfrac{f^*(x)-f^*(\widehat{x})}{\|x-\widehat{x}\|_2} + \Gamma_\Delta.$$

We can observe that, as expected, the maximum rate of variation of f_Δ is constant and equal to Γ_Δ for any $\hat{x} \in \mathcal{X}$. Instead, the maximum rate of variation of f^o is Γ_Δ plus a quantity that depends locally on the point $\hat{x}$. For this reason, when $f^* = 0$ and thus $f_\Delta = f^o$, the approach is called global Set Membership approach *which was discussed in section 1.3. Otherwise, the approach is called* local Set Membership approach.

Under the above assumptions and, in particular, under (1.19) and (1.7), we have that $f^o \in FFS$, where FFS is the Feasible Function Set defined as follows.

Definition 7. *The* Feasible Function Set *is*

$$FFS \doteq \{f : f = f^* + f_\Delta,\ f_\Delta \in \mathcal{F}(\Gamma_\Delta), \|\widetilde{y} - f(\widetilde{x})\|_q \leqslant \mu\}$$

where $\widetilde{y} = (\widetilde{y}_1, \ldots, \widetilde{y}_L)$ and $f(\widetilde{x}) \doteq (f(\widetilde{x}_1), \ldots, f(\widetilde{x}_L))$. □

According to this definition, FFS is the set of all functions consistent with the prior assumptions and data. In the Set Membership framework, the validation of prior assumptions is a fundamental step. It is usual to introduce the concept of prior assumption validation as consistency with the available data: the prior assumptions are considered validated if at least one estimate consistent with these assumptions and the data exists, i.e. if FFS is not empty, see e.g. [21] and [23].

Definition 8. *The prior assumptions are validated if $FFS \neq \varnothing$.* □

The following theorem gives a necessary and sufficient condition for the validation of prior assumptions.

Theorem 4. *$FFS \neq \varnothing$ if and only if the optimization problem (1.30) is feasible.*

Proof. See [18]. □

The following theorem shows that the approximation f^* in (1.17) is interpolatory (and thus almost-optimal). The theorem also provides an explicit bound on the worst-case approximation error. Let us define the following functions:

$$\begin{aligned}\overline{f}_\Delta(x,\widehat{\varepsilon}) &\doteq \min_{k=1,\ldots,L}\left(\delta_k(a^*)+\widehat{\varepsilon}_k+\Gamma_\Delta\|x-\widetilde{x}_k\|_2\right)\\ \underline{f}_\Delta(x,\widehat{\varepsilon}) &\doteq \max_{k=1,\ldots,L}\left(\delta_k(a^*)-\widehat{\varepsilon}_k-\Gamma_\Delta\|x-\widetilde{x}_k\|_2\right)\end{aligned} \tag{1.20}$$

$$\begin{aligned}\overline{f}(x,\varepsilon) &= f^*(x)+\overline{f}_\Delta(x,\varepsilon)\\ \underline{f}(x,\varepsilon) &= f^*(x)+\underline{f}_\Delta(x,\varepsilon)\end{aligned} \tag{1.21}$$

where $\delta_k(a^*) = \widetilde{y}_k - f^*(\widetilde{x}_k)$ (see (1.31)) and $\widehat{\varepsilon}_k \geqslant 0$, $k=1,\ldots,L$.

Theorem 5. *Assume that:*
(i) The noise affecting the measurements $\{\widetilde{x}_k,\widetilde{y}_k\}_{k=1}^L$ is bounded according to (1.7).
(ii) The function f^o is Lipschitz continuous according to (1.11).
Then, for $q=2,\infty$ and for any $p\in[1,\infty]$:
(i) The approximation f^* defined in (1.17) is interpolatory (and thus almost-optimal).
(ii) The worst-case approximation error of f^* is bounded as

$$EN(f^*) \leqslant \max_{\substack{\|\widehat{\varepsilon}\|_q\leqslant\mu\\ \|\widehat{\epsilon}\|_q\leqslant\mu}} \left\|\overline{f}_\Delta(\cdot,\widehat{\varepsilon})-\underline{f}_\Delta(\cdot,\widehat{\epsilon})\right\|_p = 2\inf_{\widehat{f}} EN(\widehat{f}). \tag{1.22}$$

Proof. See [18]. □

Suppose that $\widehat{\varepsilon}^* = \widehat{\epsilon}^* = \varepsilon$, where $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_L)$ is obtained as described in Section 1.4.1. Now let the function f^c be defined as

$$f^c(x) \doteq f^*(x) + \frac{1}{2}\left[\overline{f}_\Delta(x,\varepsilon)+\underline{f}_\Delta(x,\varepsilon)\right]. \tag{1.23}$$

where $\overline{f}_\Delta, \underline{f}_\Delta$ are given in (1.20). The following result shows that this function f^c is an optimal approximation of f^o.

Theorem 6. *Assume that:*
(i) The noise affecting the measurements $\{\widetilde{x}_k,\widetilde{y}_k\}_{k=1}^L$ is bounded according to (1.7).
(ii) The function f^o is Lipschitz continuous according to (1.11).
Then, for $q=2,\infty$ and for any $p\in[1,\infty]$:
(i) The approximation f^c defined in (1.23) is optimal.
(ii) The worst-case approximation error of f^* is given by

$$EN(f^c) = \frac{1}{2}\left\|\overline{f}_\Delta(\cdot,\varepsilon)-\underline{f}_\Delta(\cdot,\varepsilon)\right\|_p = \inf_{\widehat{f}} EN(\widehat{f}). \tag{1.24}$$

Proof. See [19]. □

In summary, we have shown that the function $f^*(x)$ is an almost-optimal approximation, whereas the function $f^c(x) \doteq f^*(x) + [\overline{f}_\Delta(x,\varepsilon) + \underline{f}_\Delta(x,\varepsilon)]/2$ is an optimal approximation of f^o. The correction term $[\overline{f}_\Delta(x,\varepsilon) + \underline{f}_\Delta(x,\varepsilon)]/2$ can be useful to check how close is f^* to the optimum: The two approximations can be be evaluated off-line on a set of data not used for identification. If the errors of the two approximations on these data are similar, it can be concluded that f^* is practically optimal. Otherwise, this comparison allows us to quantify the suboptimality level of f^* with respect to f^c.

Remark 5. *A subcase of the general theory presented here is when $f^* = 0$, in which we have the so-called* global Set Membership approach*, previously presented. Otherwise, if $f^* \neq 0$, we have the so-called* local Set Membership approach*. See also Remark 4 for an explanation of this terminology.*

1.4.1 Interval estimates

Interval estimate on the unknown function f^o are now derived. A general formulation is developed, allowing us to deal with the cases where the noise is bounded in ℓ_2 or ℓ_∞ norm (i.e. $q = 2$ or $q = \infty$ in (1.7)).

1.4.1.1 Noise bounded in ℓ_2 norm

Theorem 5 does not allow the evaluation of interval estimates, since the assumption that the noise sequence $d = (d_1, d_2, \ldots, d_L)$ is bounded in ℓ_2 norm gives no information on how the single elements d_k are bounded. In order to overcome this issue, some additional assumption has to be made on the element-wise boundedness of the noise sequence d. This kind of assumption can be obtained as follows.

Since f^* is an almost-optimal approximation of f^o, we have that $f^*(\widetilde{x}_k) \cong f^o(\widetilde{x}_k)$ and, consequently, that $d_k = \widetilde{y}_k - f^o(\widetilde{x}_k) \simeq \widetilde{y}_k - f^*(\widetilde{x}_k) \doteq \delta_k(a^*)$. It is then reasonable to assume the following relative plus absolute error bound:

$$|d_k| \leqslant \varepsilon_k \doteq \varepsilon^r |\delta_k(a^*)| + \varepsilon^a, \quad k = 1, \ldots, L \tag{1.25}$$

where the term $\varepsilon^r |\delta_k(a^*)|$ accounts for the fact that $d_k \simeq \delta_k(a^*)$ and ε^a accounts for the fact that d_k and $\delta_k(a^*)$ are not exactly equal. The parameters $\varepsilon^r, \varepsilon^a \geqslant 0$ have to be taken such that $\varepsilon^r \mu + \varepsilon^a \sqrt{L} \leqslant \mu$. Indeed, if this inequality is satisfied, (1.25) is consistent with (1.7): $\|d\|_2 \leqslant \varepsilon^r \mu + \varepsilon^a \sqrt{L} \leqslant \mu$. Following this indication, ε^r and ε^a (together with Γ_Δ) can be chosen by means of the procedure presented in [19].

In order to satisfy the assumption (1.25), the following additional constraints have to be inserted in Algorithm 1 (in particular, in (1.30) and on line 5 of (1.32)):

$$|\widetilde{y}_k - \Phi_k(\widetilde{x}_k)\, a| \leqslant \varepsilon_k, \quad k = 1, \ldots, L \tag{1.26}$$

where $\Phi_k(\widetilde{x}_k)$ is the kth row of the matrix Φ.

□

1.4.1.2 Noise bounded in ℓ_∞ norm

If the noise is bounded in ℓ_∞ norm, the required interval estimates can be obtained without introducing any further assumption. Indeed, in this case,

$$|d_k| \leqslant \varepsilon_k \doteq \mu, \quad k = 1, \ldots, L. \tag{1.27}$$

□

The following theorem, holding for both the ℓ_2 and ℓ_∞ cases, provides tight point-wise interval estimates for $f^o(x)$ and gives an expression of the worst-case error bound computable for any dimension n_x (a computationally tractable algorithm for this computation is presented in [27]).

Theorem 7. *Assume that:*
(i) The noise affecting the measurements $\{\widetilde{x}_k, \widetilde{y}_k\}_{k=1}^{L}$ *is bounded according to (1.25) or (1.27).*
(ii) The function f^o *is Lipschitz continuous according to (1.11).*
Then, for $q = 2, \infty$ *and for any* $p \in [1, \infty]$*:*
(i) The worst-case approximation error of f^* *is bounded as*

$$EN(f^*) \leqslant \left\| \overline{f}_\Delta(\cdot, \varepsilon) - \underline{f}_\Delta(\cdot, \varepsilon) \right\|_p = 2 \inf_{\widehat{f}} EN(\widehat{f}).$$

(ii) For any $x \in \mathcal{X}$, $f^o(x)$ *is tightly bounded as*

$$\underline{f}(x, \varepsilon) \leqslant f^o(x) \leqslant \overline{f}(x, \varepsilon). \tag{1.28}$$

Proof. See [18]. □

1.4.2 *Local Approach - identification algorithms*

In this section, two algorithms for the identification of the parameters a_i in (1.17) are proposed.

In order to ensure suitable regularity properties of the approximation, limiting well-known issues such as overfitting and the curse of dimensionality, we require the vector $a = (a_1, a_2, \ldots, a_N) \in \mathbb{R}^N$ of the coefficients in (1.17) to be sparse. Hence, under assumption (1.7), a solution to the identification Problem 1 could be found by solving the following optimization problem:

$$\begin{aligned} & a^0 = \arg\min_{a \in \mathbb{R}^N} \|a\|_0 \\ & \text{subject to} \quad \|\widetilde{y} - \Phi a\|_q \leqslant \mu. \end{aligned} \tag{1.29}$$

where

$$\begin{aligned} \widetilde{y} &\doteq (\widetilde{y}_1, \ldots, \widetilde{y}_L) \\ \Phi &\doteq \begin{bmatrix} \phi_1(\widetilde{x}_1) & \cdots & \phi_N(\widetilde{x}_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\widetilde{x}_L) & \cdots & \phi_N(\widetilde{x}_L) \end{bmatrix} \\ &= \begin{bmatrix} \phi_1(\widetilde{x}) & \cdots & \phi_N(\widetilde{x}) \end{bmatrix}, \end{aligned}$$

$\phi_1(\widetilde{x}) \doteq (\phi_1(\widetilde{x}_1), \ldots, \phi_1(\widetilde{x}_L))$, and $\|a\|_0$ is the ℓ_0 quasi-norm of a, defined as the number of non-zero components of a. In fact, minimizing the ℓ_0 quasi-norm of a vector corresponds to minimizing the number of its non-zero elements, i.e. to maximizing its sparsity. On the other hand, the constraint $\|\widetilde{y}-\Phi a\|_q \leqslant \mu$ ensures that the identified coefficient vector is consistent with the measured data (1.6) and the prior assumption on noise (1.7).

However, the optimization problem (1.29) cannot be easily solved, since the ℓ_0 quasi-norm is a non-convex function and its minimization is an NP-hard problem. The classical approach to overcome this issue is to replace the ℓ_0 quasi-norm with its convex envelope, i.e. the ℓ_1 norm, see e.g. [28], [29], [30]. The identification Problem 1 can thus be solved efficiently by means of the following convex optimization problem.

Algorithm 1. *Function identification 1*

$$\begin{aligned} & a^* = \arg\min_{a\in\mathbb{R}^N} \|a\|_1 \\ & \textit{subject to} \quad \|\widetilde{y}-\Phi a\|_q \leqslant \mu \end{aligned} \tag{1.30}$$

where μ can be chosen as a positive number slightly larger than μ^{min}, the minimum value for which the problem is feasible (this choice is theoretically motivated by the validation Theorem 4). Provided that $\mu > \mu^{min}$, the value of μ can be tuned to suitably manage the trade-off between accuracy and sparsity. □

Another interesting ℓ_1 algorithm, completely based on convex optimization, is now presented. As discussed below, this algorithm provides sparser solutions with respect to the standard algorithm (1.30), which is based on simple ℓ_1-norm minimization.

Without loss of generality, assume that the columns of Φ are normalized: $\|\phi_i(\widetilde{x})\|_2 = 1$, $i = 1,2,\ldots,N$. Define the following quantity:

$$\xi(a) \doteq \frac{|\delta(a)|_1 + |\delta(a)|_{K_0}}{\underline{\sigma}^2(\Phi)}$$

where $K_0 \doteq 2\|a\|_0$, $\underline{\sigma}(\Phi)$ is the minimum non-zero singular value of Φ and

$$\begin{aligned} & \delta(a) \doteq \widetilde{y}-\Phi a \\ & |w|_K \doteq \sqrt{\sum_{i\in I_K} \left(w^T\phi_i(\widetilde{x})\right)^2}, \end{aligned} \tag{1.31}$$

being I_K the set of the K largest inner products $\left|w^T\phi_i(\widetilde{x})\right|$. Let card$(\cdot)$ denote the set cardinality.

Algorithm 2. *Function identification 2*

1. *Solve the optimization problem* (1.30) *and set* $a^1 := a^*$.
2. *Let* $r(a^1) \doteq \left\{i_1,\ldots,i_j : \xi\left(a^1\right) > |a^1_{i_1}| \geqslant \ldots \geqslant |a^1_{i_j}|\right\}$ *and let* $r_\lambda(a^1)$ *denote the subset of* $r(a^1)$ *with indices in* λ. *Compute the coefficient vector* a^* *as follows:*

$$
\begin{array}{l}
\textit{for } k = 1 : card(r(a^1)) \\
\qquad c^k = \arg\min_{a \in \mathbb{R}^N} \|\widetilde{y} - \Phi a\|_q \\
\qquad \textit{subject to} \quad a_i = 0,\ \forall i \in r_\lambda(a^1) \\
\qquad \qquad \qquad \quad \lambda = \{k, \ldots, card(r(a^1))\} \\
\qquad \textit{if } \left\|\widetilde{y} - \Phi c^k\right\|_q \leqslant \mu \\
\qquad \qquad a^* := c^k \\
\qquad \qquad \textit{break} \\
\qquad \textit{end} \\
\textit{end}
\end{array}
\tag{1.32}
$$

□

The rationale behind Algorithm 2 can be explained as follows: In step 1., an optimization problem similar to (1.29) is solved, where the ℓ_0 quasi-norm is replaced by the ℓ_1 norm. The ℓ_1 norm is the convex envelope of the ℓ_0 quasi-norm, and its minimization yields a sparse vector a^1 [28], [29], [30]. However, it is not guaranteed that all the non-zero elements of a^1 are necessary to have $\|\widetilde{y} - \Phi a^1\|_q \leqslant \mu$. In step 2., only the elements of a^1 larger than $\xi\left(a^1\right)$ are kept (indeed, $\xi\left(a^1\right)$ discriminates between "important" and "less important" vector components, [22]), while the remaining ones, ordered by decreasing amplitude, are progressively included to form the vector a^*. The algorithm stops when $\|\widetilde{y} - \Phi c^k\|_q \leqslant \mu$. The solution provided by step 2. is thus a vector a^* where the number of non-zero elements is further reduced with respect to the initial sparse solution a^1. The sparse approximation is given by (1.17), with $a_i = a_i^*$, $\forall i$.

1.5 Nonlinear Set Membership Identification : Quasi-Local Approach

In this section, the so-called quasi-local Nonlinear Set Membership approach is presented, capturing the advantages of the global and local approaches of Section 1.3, 1.4 and avoiding some drawbacks of these two. On one hand, the quasi-local approach allows the derivation of significantly less conservative uncertainty bounds with respect to the global approach of Section 1.3. On the other hand, the quasi-local approach does not require to choose a suitable parametric form for the filter, as done in the local approach of Section 1.4. The filter is obtained directly from the data in a non-parametric closed form.

Based on assumption (1.11), we can define the following quantity, called the *quasi-local Lipschitz parameter*:

$$
\gamma(x) \doteq \sup_{\widehat{x} \in \mathcal{X}, \widehat{x} \neq x} \frac{|f_o(x) - f_o(\widehat{x})|}{\|x - \widehat{x}\|_2} \tag{1.33}
$$

Obviously, the Lipschitz constant of f on $\mathcal{X}$ is given by

$$\Gamma = \sup_{x \in \mathcal{X}} \gamma(x). \tag{1.34}$$

Lemma 1. *For any $x \in \mathcal{X}$, a $\gamma(x)$ exists, such that*

$$|f_o(x) - f_o(\widehat{x})| \leqslant \gamma(x) \, \|x - \widehat{x}\|_2 \, , \, \forall \widehat{x} \in \mathcal{X}.$$

Proof. The statement follows directly from (1.33). □

Let us now suppose that the quasi-local Lipschitz parameters $\gamma(\widetilde{x}_k)$, $k = 1, ..., L$, are known or can be estimated (a method for performing such an estimation is given in Section 1.6.1). Assume also that the noise is bounded in ℓ_∞ norm according to (1.7), with $q = \infty$ and $\varepsilon \doteq \mu$. On the basis of this information, we can define the following function set:

$$\mathcal{F}_{ql} \doteq \{f : |f(x) - f(\widehat{x}_k)| \leqslant \gamma(\widehat{x}_k) \, \|x - \widehat{x}_k\|_2 \, , \forall x \in \mathcal{X}, \, k = 1, \ldots, L\}. \tag{1.35}$$

This allows us to define the *Feasible Function Set* as follows.

Definition 9. *The* Feasible Function Set *is*

$$FFS \doteq \left\{ f \in \mathcal{F}_{ql} : \|\widetilde{y} - f\,(\widetilde{x})\|_q \leqslant \mu \right\} \tag{1.36}$$

where $\widetilde{y} = (\widetilde{y}_1, \ldots, \widetilde{y}_L)$ *and* $f\,(\widetilde{x}) \doteq (f\,(\widetilde{x}_1), \ldots, f\,(\widetilde{x}_L))$. □

As discussed in Section 1.4, FFS is the set of all functions consistent with the prior assumptions and data. The prior assumptions are considered validated if at least one estimate consistent with these assumptions and the data exists, i.e. if FFS is not empty, see e.g. [21] and [23].

Definition 10. *The prior assumptions are validated if* $FFS \neq \varnothing$. □

The following theorem gives a necessary and a sufficient condition for the validation of prior assumptions. Define the functions

$$\begin{aligned} \overline{f}(x) &\doteq \min_{k=1,\ldots,L} (\overline{h}_k + \gamma(\widetilde{x}_k) \, \|x - \widetilde{x}_k\|_2) \\ \underline{f}(x) &\doteq \max_{k=1,\ldots,L} (\underline{h}_k - \gamma(\widetilde{x}_k) \, \|x - \widetilde{x}_k\|_2) \end{aligned} \tag{1.37}$$

where $\overline{h}_k \doteq \widetilde{y}_k + \varepsilon_k$ and $\underline{h}_k \doteq \widetilde{y}_k - \varepsilon_k$.

Theorem 8. *(i) A necessary condition for prior assumptions to be validated is* $\overline{f}(\widetilde{x}_k) \geqslant \underline{h}_k \, , \underline{f}(\widetilde{x}_k) \leqslant \overline{h}_k$, $k = 1, ..., L$.
(ii) A sufficient condition for prior assumptions to be validated is $\overline{f}(\widetilde{x}_k) > \underline{h}_k \, , \underline{f}(\widetilde{x}_k) < \overline{h}_k$, $k = 1, ..., L$.

Proof. see [31]. □

Let us now define the function

$$f^c(x) \doteq \frac{1}{2} [\underline{f}(x) + \overline{f}(x)] \tag{1.38}$$

where $\underline{f}(x)$ and $\overline{f}(x)$ are given in (1.37). The next result shows that the approximation f^c is optimal for any L_p norm. The theorem also provides an explicit bound on the worst-case approximation error.

Theorem 9. *Assume that:*
(i) The noise affecting the measurements $\{\widetilde{x}_k, \widetilde{y}_k\}_{k=1}^L$ *is bounded according to (1.7).*
(ii) The function f^o *is Lipschitz continuous according to (1.11).*
Then, for $q = \infty$ *and for any* $p \in [1, \infty]$*:*
(i) The approximation f^c *defined in* (1.38) *is optimal.*
(ii) The worst-case approximation error of f^c *is bounded as*

$$EN(f^c) = \frac{1}{2} \left\| \overline{f} - \underline{f} \right\|_p = \inf_{\widehat{f}} EN(\widehat{f}) \doteq \mathcal{R}_{\mathcal{I}}. \tag{1.39}$$

Proof. see [31]. □

1.5.1 Interval estimates

The interval estimates of the unknown function f^o are now given for the cases where the noise is bounded in ℓ_∞ norm (i.e. $q = \infty$ in (1.7)).

Theorem 10. *Assume that:*
(i) The noise affecting the measurements $\{\widetilde{x}_k, \widetilde{y}_k\}_{k=1}^L$ *is bounded according to (1.7).*
(ii) The function f^o *is Lipschitz continuous according to (1.11).*
Then, for $q = \infty$ *and for any* $x \in \mathcal{X}$, $f^o(x)$ *is tightly bounded as*

$$\underline{f}(x) \leqslant f^o(x) \leqslant \overline{f}(x). \tag{1.40}$$

Proof. see [18]. □

Remark 6. *The point-wise bounds (1.40) provide an interval estimate of the unknown value* $f^o(x)$*. Interval estimates allow us to quantify the uncertainty associated with the identification process, and are thus important in system and control applications. Indeed, these estimates can be used e.g. for robust control design, [32], [33], prediction interval evaluation, [34], and fault detection, see the following sections.*

Remark 7. *A study on the relation between the representativeness and length of the available data and the quality of the approximation can be found in [27]. This study relies on the computation of the radius of information* $\mathcal{R}_{\mathcal{I}}$ *(see Definition 2), representing the minimum worst-case error that can be achieved from the available prior and experimental information. Based on this concept, a methodology is proposed in [27], where* $\mathcal{R}_{\mathcal{I}}$ *is used to evaluate the "level of information" provided by a given data set and to obtain precise indications on the quality of the approximation that can be obtained.*

Fig 1.1 shows the comparison between the global and quasi-local Set Membership bounds for a nonlinear function. In figure 1a, a global lipschitz constant Γ was assumed for the function f^o. In figure 1b, a quasi-local lipschitz parameter $\gamma(x)$ was assumed. It can be noted that the resulting uncertainty bounds are clearly tighter in the quasi-local case, especially in regions where the function is relatively "flat".

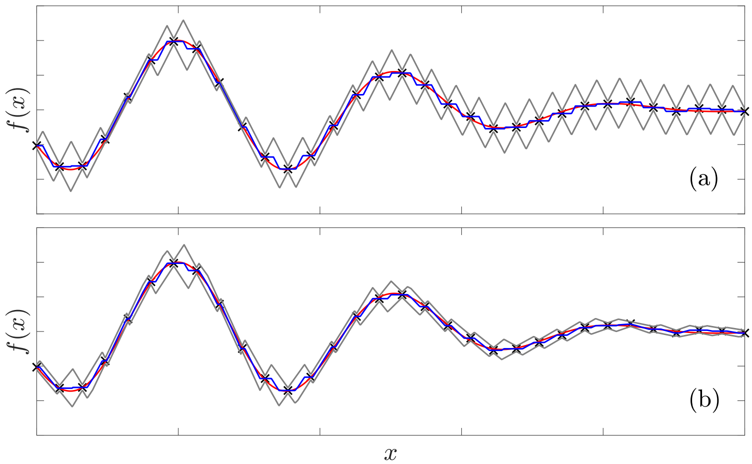

Figure 1.1 *(a) global bounds; (b) quasi-local bounds; $f^o(x)$: red line; measurements: black cross; $f^c(x)$: blue line; $\overline{f}(x), \underline{f}(x)$: grey line.*

1.6 Parameter Estimation and Adaptive Set Membership Model

1.6.1 Parameter Estimation

Estimates of the noise bound μ, Lipschitz constants Γ, Γ_Δ, and the quasi-local lipschitz parameter $\gamma(x)$ such that the assumptions are validated can be obtained by means of two algorithms given in [35] and reported in the following. The first algorithm is directly taken from [35], while the second one is a generalization of the corresponding one in [35].

Algorithm 3. *noise bound estimation* $\widehat{\mu}$

1. *Choose a "small"* $\rho > 0$. *for example:*
   $\rho = 0.01 max_{k,t=1,...,L-1} \|\widetilde{x}_k - \widetilde{x}_t\|$.
2. *Find the set of indexes:* $I_k \doteq \{k : \|\widetilde{x}_k - \widetilde{x}_t\| \leqslant \rho\}$. *if* $I_k = \emptyset$ *for all* $k = 1,...,L-1$, *go to step 1 and choose a larger* ρ.
3. *For* $k = 1,...,L-1$ *compute* $\delta\widetilde{y}_k = \max_{i \in I_k} |\widetilde{y}_k - \widetilde{y}_i|$. *If* $I_k = \emptyset$, *set* $\delta\widetilde{y}_k = \infty$.
4. *Obtain the estimate* $\widehat{\mu}$ *of the noise bound* μ *as*
   $\widehat{\mu} = \frac{1}{2N} \sum_{k \in Q} \delta\widetilde{y}_k$
   where $Q \doteq \{k \in \{1,...,L-1\} : \delta\widetilde{y}_k < \infty\}$ *and* $N \doteq \mathrm{card}(Q)$

□

Algorithm 4. *Lipschitz parameter estimation* $\widehat{\Gamma}$, $\widehat{\Gamma}_\Delta$, $\widehat{\gamma}(x)$

1. *For $t,k = 1,...,L-1$ and $\widetilde{x}_k \neq \widetilde{x}_t$, compute*

$$\widehat{\Gamma} = \max_{k,t=1,...,L-1} \begin{cases} \frac{|\widetilde{y}_k - \widetilde{y}_t| - 2\widehat{\mu}}{\|\widetilde{x}_k - \widetilde{x}_t\|_2} & \text{if} \quad |\widetilde{y}_k - \widetilde{y}_t| > 2\widehat{\mu} \\ 0 & \text{otherwise.} \end{cases} \tag{1.41}$$

2. *For $t,k = 1,...,L-1$ and $\widetilde{x}_k \neq \widetilde{x}_t$, compute*

$$\widehat{\Gamma}_\Delta = \max_{k,t=1,...,L-1} \begin{cases} \frac{\left|\widetilde{\delta}_k - \widetilde{\delta}_t\right| - 2\widehat{\mu}}{\|\widetilde{x}_k - \widetilde{x}_t\|_2} & \text{if} \quad |\widetilde{\delta}_k - \widetilde{\delta}_t| > 2\widehat{\mu} \\ 0 & \text{otherwise.} \end{cases} \tag{1.42}$$

3. *For $k = 1,...,L-1$ and $\widetilde{x}_k \neq \widetilde{x}_t$, compute*

$$\widehat{\gamma}(\widetilde{x}_k) = \max_{t=1,...,L-1} \begin{cases} \frac{|\widetilde{y}_k - \widetilde{y}_t| + 2\widehat{\mu}}{\|\widetilde{x}_k - \widetilde{x}_t\|_2} & \text{if} \quad |\widetilde{y}_k - \widetilde{y}_t| > 2\widehat{\mu} \\ 0 & \text{otherwise.} \end{cases} \tag{1.43}$$

□

The following theorems show that, under reasonable density conditions on the noise, the estimates given by these two algorithms converge to the corresponding true values.

Theorem 11. *(Theorem 2 of [35]) Let the set $\{\widetilde{x}_k, d_k\}_{k=1}^{L}$ appearing in (1.2) be dense on $\mathcal{X} \times B_\varepsilon$ as $L \to \infty$. Then,*

$$\lim_{L\to\infty} \widehat{\mu} = \mu. \qquad \square$$

Theorem 12. *(Theorem 3 of [35]) Let the set $\{\widetilde{x}_k, d_k\}_{k=1}^{L}$ appearing in (1.2) be dense on $\mathcal{X} \times B_\varepsilon$ as $L \to \infty$. Then,*

$$\lim_{L\to\infty} \widehat{\Gamma} = \Gamma. \qquad \square$$

1.6.2 Adaptive Set Membership Model

In many applications, it may happen that the dynamics of the system changes over time or the model is not accurate enough in the whole domain of $\mathcal{X}$. One of the advantages of nonlinear Set Membership (global and quasi-local approach) is that it can be easily made adaptive since no optimization problem needs to be solved online.

As discussed in section 1.2, in the set membership framework, the model accuracy is defined by the radius of information and it can be computed in a deterministic way [27]. In the global and quasi-local approaches, the model is basically defined by the measurement set $\mathcal{D}$, noise bound μ and lipschitz parameters Γ or $\gamma(x)$. Therefore, the model can be made adaptive by updating online the measurement set and the lipschitz parameters.

Let us define the error function as follows:

$$f^e(x) = \frac{1}{2}[\overline{f}(x) - \underline{f}(x)] \tag{1.44}$$

This function allows us to write the radius of information as

$$\mathcal{R}_{\mathcal{I}} = \|f_e(\cdot)\|_p. \tag{1.45}$$

Suppose that a model with a desired radius of information $\mathcal{R}_d$ is looked for, where $\mathcal{R}_d < \mathcal{R}_{\mathcal{I}}$. If, at time instant t, $f^e(x_t) > \mathcal{R}_d$, it means that the model error in that point is larger than the desired radius. Therefore, the new measurement can be added to the set $\mathcal{D}$ in order to increase the model accuracy.

In the case where the dynamics of the system changes over time, a time label can be assigned to each element of the set $\mathcal{D}$ as $\mathcal{D}(L) = \{\widetilde{x}_k, \widetilde{y}_k, t_k\}_{k=1}^{L}$ where t_k indicates the time each measurement was taken. Then, since the model is running online, at each time instant t, we can eliminate the measurements which were taken at $t - \delta t$, where δt is a desired value which depends on the system dynamics variation.

Algorithm 5. *Online Update of the Set Membership Model*

1. *Define the measurement set $\mathcal{D}$ as $\mathcal{D}(L) = \{\widetilde{x}_k, \widetilde{y}_k, t_k\}_{k=1}^{L}$.*
2. *At time step t, compute the vector x_t and measure the system output y_t.*
3. *If $f^e(x_t) > \mathcal{R}_d$. Then,*

$$\begin{aligned} &\mathcal{D}(L+1) = \mathcal{D}(L) \cup \{\widetilde{x}_{L+1} = x_t, \widetilde{y}_{L+1} = y_t, t_{L+1} = t\} \\ &L = L+1. \end{aligned} \tag{1.46}$$

4. *Find the set of indexes $I_k = \{k : t_k \in \mathcal{D}, t_k < t - \delta t\}$. Then,*

$$\begin{aligned} &\mathcal{D}(L-N) = \mathcal{D}(L) \setminus \{\widetilde{x}_k, \widetilde{y}_k, t_k\}_{k \in I_k} \\ &L = L-N \end{aligned} \tag{1.47}$$

   where $N \doteq \mathrm{card}(I_k)$.
5. *If $\mathcal{D}$ is changed, update the model lipschitz parameters according to algorithm 4.*

□

1.7 Summary of Set Membership Fault Detection Procedure

The main steps of the proposed Set Membership fault detection method are now summarized.

Offline operations

1. *Define the measurement set $\mathcal{D}$ according to (1.2).*
2. *Estimate the noise bound μ according to Algorithm 3.*
3. *In the case of local approach, estimate a preliminary approximation f^* according to Algorithm 1 or 2.*
4. *Estimate the Lipschitz parameters according to Algorithm 4. In the case of global approach, compute $\widehat{\Gamma}$; in the case of quasi-local approach, compute $\widehat{\gamma}(x)$; in the case of local approach $\widehat{\Gamma}_\Delta$.*

Online operations

1. *At each time step:*

$$\begin{array}{ll} \textit{If} & \tilde{y}_k > \overline{f}(\tilde{x}_k) + \varepsilon_k \quad \textit{or} \quad \tilde{y}_k < \underline{f}(\tilde{x}_k) - \varepsilon_k \\ \textit{Then} & \textit{Fault} = 1 \\ \textit{Else} & \textit{Fault} = 0 \end{array}$$

2. *In the case of adaptive algorithm, update the Set Membership model according to Algorithm 5 if no fault has accrued in the system and the system is not recovering from a fault.*

Regrading the online operations, since the Set Membership model describes the normal dynamic behaviour of the system, it has to be updated when no other dynamics are involved. In other words, during a fault or immediately after a fault the measurements correspond to an abnormal behaviour of the system and therefore they are not suitable for updating the model.

1.8 Example: Fault Detection for a Drone Actuator

In this example, the proposed fault detection algorithms was tested on a real drone actuator, in a laboratory experimental setup. The actuator is composed of three main components: a Brushless DC motor, a driver and a propeller. The motor makes the propeller rotate and this rotation produces the required thrust and torque to drive a drone. The torque produced by the motor is given by

$$T = K_t(I - I_0) \tag{1.48}$$

where T is the motor torque, I is the input current, I_0 is the current when there is no load on the motor and K_t is the torque constant of the motor. The voltage across the motor is given by

$$V = R_m I + L_m \frac{\mathrm{d}I}{\mathrm{d}t} + K_e \omega \tag{1.49}$$

where V is the voltage drop across the motor, R_m and L_m are the motor resistance and inductance, respectively, and K_e is the motor speed constant. The motor is attached to a propeller, and therefore we have

$$T = B_m \omega + (J_m + Jp) \frac{\mathrm{d}\omega}{\mathrm{d}t} + T_D \tag{1.50}$$

where B_m is the motor friction, ω is the motor angular speed, J_m and J_p are the motor and propeller moments of inertia, respectively, and T_D is the propeller drag torque, given by

$$T_D = C_p \rho D^5 \omega^2 \tag{1.51}$$

where C_p is the propeller power coefficient, ρ is air density and D is the propeller diameter. The power coefficient C_p is a nonlinear function of the propeller speed, which is difficult to derive from the geometric shape of the propeller [36].

1.8.1 Experimental Setup

A brushless DC motor (RIMFIRE.10) attached to a propeller (APC 10x4.7p) is connected to a driver and a 200 watt power supply. The input command of the driver is a PWM signal with a duly cycle proportional to the voltage across the motor. An encoder is attached to the back of the motor, to measure the angular speed of the propeller. The input command signal and the encoder are connected to a PC running MATLAB through a National Instrument data acquisition device (PCI-6289). Data acquisition and online fault detection were carried out using the SIMULINK Desktop Real-Time. The drone actuator system is shown in Figure 1.2.

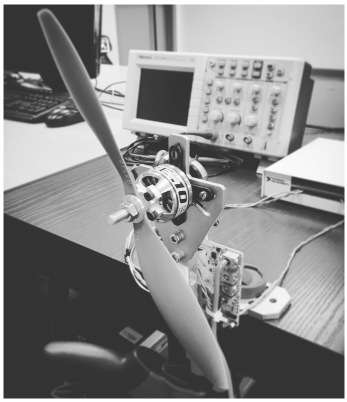

Figure 1.2 Drone Actuator

1.8.2 Nonlinear Set Membership Fault Detection

In order to identify a model, an experiment was carried out, where an amplitude modulated pseudo random binary sequence (APRBS) command input with a duration of 100 seconds was applied to the actuator. From this experiment, a set of data was collected, using a sampling time of $T_s = 0.05s$. The dataset was divided into an identification set, composed by the first 1000 data and a validation set, composed by the remaining 1000 data. See Figure 1.4 (a). The measurement set was defined according to (1.2) as follows:

$$\begin{aligned} \mathcal{D} &= \{\widetilde{x}_k, \widetilde{y}_k\}_{k=4}^{1000} \\ \widetilde{x}_k &= [\widetilde{y}_{k-1} \quad \widetilde{y}_{k-2} \quad \widetilde{u}_{k-2} \quad \widetilde{u}_{k-3}] \end{aligned} \tag{1.52}$$

To evaluate the performance of the fault detection algorithm, a second experiment was carried out using another APRBS input command that was not used for identification nor validation. This second experiment had a duration of 30 seconds and during this experiment, in order to introduce a fault scenario, a sheet of paper was placed between the blades, approximately every 3 seconds, for a total of 8 times. In Figure 1.3 (b), the fault occurrences are denoted by the circles. During this experiment, the following fault detection algorithms were running online.

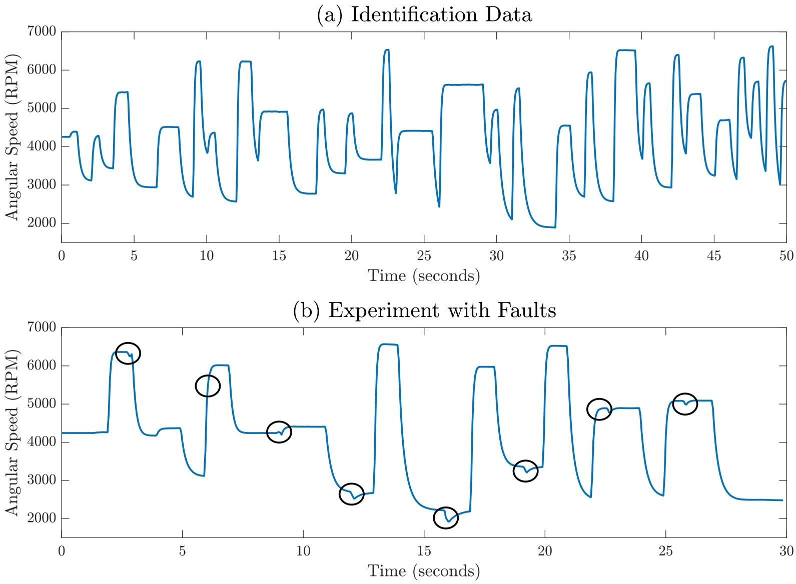

Figure 1.3 *(a) First experiment used for identification, (b) Second experiment online fault detection test (Black circles are where the faults has occurred)*

1.8.2.1 Global Approach

A nonlinear set membership model was obtained assuming a global constant bound on the function gradient according to section 1.3. A constant lipschitz parameter $\Gamma = 0.7$ and a noise bound $\mu = 15$ were estimated according to Algorithms 3 and 4. The model was tested in simulation on the validation set and the root mean square simulation error (RMSE) was 42 RPM. Then, the fault detection algorithm was applied online, as discussed above. Figure 1.4 shows the model intervals and the bound violations. The model was able to detect 7 out of 8 faults.

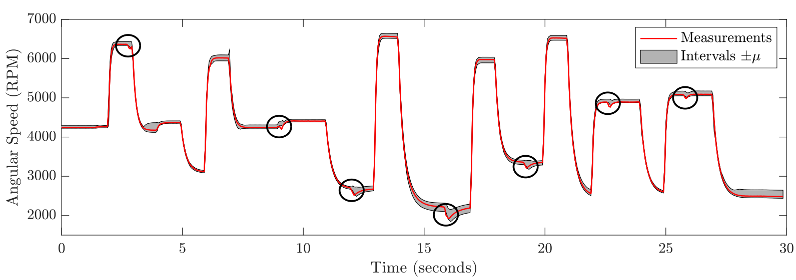

Figure 1.4 *Global Approach; Black lines, $\overline{f}+\mu, \underline{f}-\mu$. Red line, $\widetilde{y}$. Circles, bound violation*

1.8.2.2 Quasi-Local Approach

A nonlinear set membership quasi-local model was obtained according to section 1.5. A quasi-local lipschitz parameter γ was derived according to Algorithm 4. The model was tested in simulation on the validation set and the RMSE was 30 RPM. Then, the fault detection algorithm was applied online, as discussed above. Figure 1.5 shows the model intervals and the bound violations. The model was able to detect all the 8 faults.

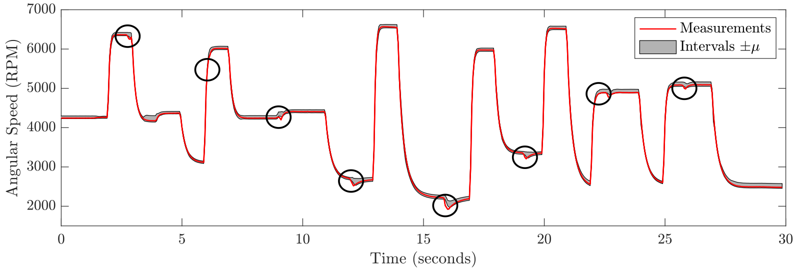

Figure 1.5 Quasi-Local Approach; Black lines, $\overline{f}+\mu, \underline{f}-\mu$. Red line, $\widetilde{y}$. Circles, bound violation

1.8.2.3 Local Approach

A preliminary approximation f^* was derived according to Algorithm 1, using polynomial basis functions up to degree 5. No improvements were observed considering higher degrees. The number of basis functions in (1.30) is 81 where, due to ℓ_1 sparsification, only 33 have a non-null coefficient. The nonlinear set membership local model was obtained according to section 1.4. A lipschitz parameter $\Gamma_\Delta = 0.2$ was derived according to Algorithm 4. The model was tested in simulation on the validation set and the RMSE was 23 RPM. Then, the fault detection algorithm was applied online, as discussed above. Figure 1.5 shows the model intervals and the bound violations. The model was able to detect all 8 faults.

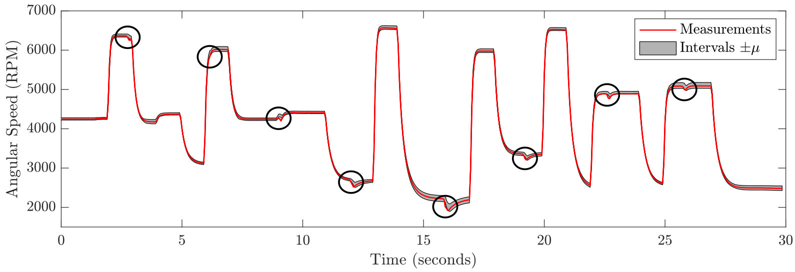

Figure 1.6 Local Approach; Black lines, $\overline{f}+\mu, \underline{f}-\mu$. Red line, $\widetilde{y}$. Circles, bound violation

1.9 Conclusions

A novel fault detection approach, based on Set Membership interval estimates, has been presented, allowing us to overcome several problems of the standard techniques. Its effectiveness has been demonstrated in a laboratory study, related to fault detection for a real propeller. The possibility of extending the approach to fault isolation is currently under investigation.

References

[1] Zhang X, Polycarpou MM, Parisini T. A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. IEEE Transactions on Automatic Control. 2002;47(4):576–593.

[2] Patton RJ, Frank PM. Fault diagnosis in dynamic systems: theory and applications. Prentice Hall; 1989.

[3] Gertler J. Fault Detection and Diagnosis in Engineering Systems. New York: Marcel Dekker; 1998.

[4] Persis CD, Isidori A. A geometric approach to nonlinear fault detection and isolation. IEEE Transactions on Automatic Control. 2001;46(6):853–865.

[5] Isermann R. Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance. Berlin, Germany: Springer; 2006.

[6] Ferrari RMG, Parisini T, Polycarpou MM. Distributed Fault Detection and Isolation of Large-Scale Discrete-Time Nonlinear Systems: An Adaptive Approximation Approach. IEEE Transactions on Automatic Control. 2012;57(2):275–290.

[7] Liu B, Si J. Fault isolation filter design for linear time-invariant systems. IEEE Transactions on Automatic Control. 1997;42:704–707.

[8] Simani S, Fantuzzi C, Patton RJ. Model-based fault diagnosis in dynamic systems using identification techniques. Springer Science & Business Media; 2013.

[9] Chen J, Patton RJ. Robust model-based fault diagnosis for dynamic systems. vol. 3. Springer Science & Business Media; 2012.

[10] Patton RJ, Frank PM, Clark RN. Issues of fault diagnosis for dynamic systems. Springer Science & Business Media; 2013.

[11] Armengol J, Trave-Massuyes L, Vehi J, et al. A survey on interval model simulators and their properties related to fault detection. Annual Reviews in Control. 2000;24:31–39.

[12] Puig V, Quevedo J, Escobet T, et al. Passive Robust Fault Detection of Dynamic Processes Using Interval Models. IEEE Transactions on Control Systems Technology. 2008;16(5):1083–1089.

[13] Reppa V, Tzes A. Fault detection and diagnosis based on parameter set estimation. IET Control Theory and Applications. 2010;5(1):69–83.

[14] Blesa J, Puig V, Saludes J. Identification for passive robust fault detection using zonotope-based set-membership approaches. International Journal of Adaptive Control and Signal Processing. 2011;25(9):788–812.

[15] Raissi T, Videau G, Zolghadri A. Interval observer design for consistency checks of nonlinear continuous-time systems. Automatica. 2010;46(3):518–527.

[16] Tornil-Sin S, Ocampo-Martinez C, Puig V, et al. Robust fault detection of non-linear systems using set-membership state estimation based on constraint satisfaction. Engineering Applications of Artificial Intelligence. 2012;25(1):1–10.

[17] Tornil-Sin S, Ocampo-Martinez C, Puig V, et al. Robust Fault Diagnosis of Non Linear Systems Using Interval Constraint Satisfaction and Analytical Redundancy Relations. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2014;44(1):18–29.

[18] Novara C. Sparse set membership identification of nonlinear functions and application to fault detection. International Journal of Adaptive Control and Signal Processing. 2016;30(2):206–223.

[19] Milanese M, Novara C. Set Membership Identification of Nonlinear Systems. Automatica. 2004;40/6:957–975.

[20] Traub JF, Wasilkowski GW, Woźniakowski H. Information-Based Complexity. Academic Press, Inc.; 1988.

[21] Milanese M, Norton J, Lahanier HP, et al. Bounding Approaches to System Identification. Plenum Press; 1996.

[22] Novara C. Sparse identification of nonlinear functions and parametric Set Membership optimality analysis. In: American Control Conference. San Francisco, USA; 2011. .

[23] Chen J, Gu G. Control-Oriented System Identification: An H_∞ Approach. New York: John Wiley & Sons; 2000.

[24] Sjöberg J, Zhang Q, Ljung L, et al. Nonlinear Black-box Modeling in System Identification: a Unified Overview. Automatica. 1995;31:1691–1723.

[25] Hsu K, Novara C, Vincent T, et al. Parametric and nonparametric curve fitting. Automatica. 2006;42/11:1869–1873.

[26] Novara C, Vincent T, Hsu K, et al. Parametric identification of structured nonlinear systems. Automatica. 2011;47(4):711 – 721.

[27] Milanese M, Novara C. Computation of local radius of information in SM-IBC identification of nonlinear systems. Journal of Complexity. 2007;23:937–951.

[28] Fuchs JJ. Recovery of exact sparse representations in the presence of bounded noise. IEEE Transactions on Information Theory. 2005 oct;51(10):3601 – 3608.

[29] Tropp JA. Just relax: convex programming methods for identifying sparse signals in noise. IEEE Transactions on Information Theory. 2006 mar;52(3):1030 –1051.

[30] Donoho DL, Elad M, Temlyakov VN. Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Transactions on Information Theory. 2006 jan;52(1):6 – 18.

[31] Karimshoushtari M, Novara C. Design of Experiments for nonlinear dynamic systems based on Set Membership identification. Internal Report Politecnico di Torino. 2018;.

[32] Freeman A, Kokotovic V. Robust Nonlinear Control Design. Boston: Birkhuser; 1996.

[33] Qu Z. Robust Control of Nonlinear Uncertain Systems. Wiley series in nonlinear science; 1998.

[34] Milanese M, Novara C. Set Membership Prediction of Nonlinear Time Series. IEEE Transactions on Automatic Control. 2005;50(11):1655–1669.

[35] Fagiano L, Novara C. Learning a nonlinear controller from data: theory, computation and experimental results. IEEE Transactions on Automatic Control. 2016;61(7):1854–1868.

[36] Deters R, Selig M. Static testing of micro propellers. In: 26th AIAA applied aerodynamics conference; 2008. p. 6246.